UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/716,202	11/19/2003	Mark D. Wasson	P68795US0	4117

136 7590 08/18/2009

JACOBSON HOLMAN PLLC
400 SEVENTH STREET N.W.
SUITE 600
WASHINGTON, DC 20004

EXAMINER	
SHAH, PARAS D	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
08/18/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/716,202	WASSON ET AL.	
	Examiner	Art Unit	
	PARAS SHAH	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *05/14/2009*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *4,5,7,9,11,12,15,16,19-21,25,41,43,45,48,49,51,52 and 56-62* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *4, 5, 7, 9, 11, 12, 15, 16, 19-21, 25, 41, 43, 45, 48, 49, 51, 52, 56-62* _____ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This communication is in response to the Arguments and Amendments filed on 05/14/2009. Claims 4, 5, 7, 9, 11, 12, 15, 16, 19-21, 25, 41, 43, 45, 48, 49, 51, 52, 56-62 are pending and have been examined. The Applicants' amendment and remarks have been carefully considered, but they do not place the claims in condition for allowance.

2. All previous objections and rejections directed to the Applicant's disclosure and claims not discussed in this Office Action have been withdrawn by the Examiner.

Response to Arguments

3. Applicant's arguments (pages 14-27) filed on 05/14/2009 with regard to the rejections applied under 35 USC 103 have been fully considered but are moot in view of new grounds for rejection. Specifically the newly added limitations of a "client-server hardware architecture" and "tree traversal functionality based on a language that can navigate XML representations of text."

With respect to the objections to the Specification and the new matter rejections under 35 USC 112, 1st paragraph, the applicant submits a Declaration under 37 CFR 1.132 indicating that a person of ordinary skilled in the art would recognize that the feature of "in a single view of a document expressed as inline XML." The Examiner respectfully disagrees with this assertion. The Applicant provides several instances in the Specification showing that such is inherent. The Examiner. upon review of these sections, has not seen that such feature is

inherent. Section 2163.07(a) indicates the following "Inherency, however, may not be established by probabilities or possibilities. The mere fact that a certain thing may result from a given set of circumstances is not sufficient." The Examiner asserts that the various citations towards the specification are merely examples and such is not required to occur all of the time. For example, paragraphs [0074]-[0077] shows a single tree representation but it is still uncertain as to how this is a single view and what is meant by a single view. If the font size of the XML is changed then the XML can exceed a single view appearing in two different pages. Hence, since this is a possibility that may occur, such feature is not an inherent feature. For this reason, the applicants' arguments are not persuasive.

With respect to the rejections under 35 USC 103, the applicants' arguments with respect to the teachings of Simov are considered moot in light of the amended claims.

With respect to the Applicant arguments under 35 USC 103, an argument is made that Krauthammer does not specifically teach the concept of resolving conflicting annotation boundaries since Krauthammer is only for a single specific semantic components that requires knowledge about a domain. Further, the Applicants assert that Krauthammer does not address the other types of annotators from applying multiple independent grammars. In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e.,

"resolving conflicting annotation boundaries resulting from annotations produced by multiple independent annotators) are not recited in the rejected claim(s). Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). None of the claims recite the feature that is being argued by the applicants. Neither is there an inherent feature that is recited that generalizes across different types of annotators. For this reasons, the Applicants arguments are not persuasive.

With respect to claim 15 the rejections under 35 USC 103, the Applicants argue that editing functions are not fact extraction. The Examiner disagrees with this assertion. The claim is not specific as to whether such naming is done during fact extraction or after as a post processing task. Hence, the Applicants arguments are not persuasive.

With respect to claim 11 the rejections under 35 USC 103, the Applicants argue that Marcus uses notational devices to indicate whether two discontinuous constituents are related. Applicant's arguments fail to comply with 37 CFR 1.111(b) because they amount to a general allegation that the claims define a patentable invention without specifically pointing out how the language of the claims patentably distinguishes them from the references. Rather the Applicants arguments seem to be confusing and merely indicate what the reference of Marcus shows. It does not indicate how the claim distinguishes from the references. Claim 11 only recites the recognizing of non-contiguous attributes.

Marcus teaches such in page 117, sect. 6, sect paragraph and the example at bottom of page 117 (right hand column). A number is assigned to overcome the problem and therefore identifies such circumstance. Therefore, it is unclear as to what the Applicants are intending to point out in reference to the current claim. Hence, the Applicants arguments are not persuasive.

With respect to claim 20 the rejections under 35 USC 103, the Applicants have not amended the claims but rather has been amended to overcome a 35 USC 101 issue and therefore no new reference has been applied in view of the Arguments made by the Examiner presented above.

Response to Amendment

4. Applicants' amendments filed on 05/14/2009 have been fully considered. The newly amended limitations necessitate new grounds of rejection. Specifically, the newly added limitation of "client-server hardware architecture" and "tree traversal functionality based on a language that can navigate XML representations of text" necessitates new grounds for rejection.

Specification

5. The amendment filed 08/26/2008 is objected to under 35 U.S.C. 132(a) because it introduces new matter into the disclosure. 35 U.S.C. 132(a) states that no amendment shall introduce new matter into the disclosure of the invention. The added material which is not supported by the original disclosure is as follows: In the amended

Specification, paragraphs [00099] and [000115], where "FEX annotations are captured in a single view of the document expressed as inline XML" and in paragraphs [000175] and [000186], where "in a single view of the annotated document" are considered to be new matter..

Applicant is required to cancel the new matter in the reply to this Office Action.

6. The specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. See 37 CFR 1.75(d)(1) and MPEP § 608.01(o). Correction of the following is required: "machine readable storage" is not defined in the Specification. The Applicant is advised to change the terminology to "computer usage storage medium" which was included in the claims submitted on 11/19/2003 at the time of filing and to amend the Specification using the same language as in claim 26 submitted on 11/19/2003 such that no new matter is introduced.

Claim Objections

7. Claim 16 is objected to because of the following informalities: "machine readable storage" recites new terminology which was not found in the originally filed Specification and hence the scope the Applicant is intending to encompass is uncertain. The Applicant is advised to change the terminology to "computer usage storage medium" which was included in the claims submitted on 11/19/2003 at the time of filing. Appropriate correction is required.

8. Claim 16 is objected to because of the following informalities: "machine readable storage having stored thereon a computer program product application" should be

changed to "machine readable storage storing a computer readable program code, which is executed by a processor, where the computer readable program code application includes... ". Appropriate correction is required.

Claim Rejections - 35 USC § 112

9. The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

10. Claims 57-59 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. Specifically, the limitation of "annotating the text represents the annotated text as a single view of the document expressed as inline XML" has been newly added subject matter, which was not defined in the Specification as originally filed.

11. Claims 7, 12, and 20 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. Specifically, the amended claims recite "a client-server hardware

architecture." Such amendment was found to be described in the Specification paragraphs [0518]-[0521]. However, these paragraphs fail to recite how the extraction tool set using this client-server hardware architecture is interacting with structural elements of a computer system. Rather, the specification describes an operating system and the fact extraction tool as claimed being utilized by the operating system (i.e. software relationships), describing block elements within a computer. Thus, the disclosure does not provide sufficient disclosure regarding the apparatus describing the interrelationships between the software and hardware elements. See MPEP 2164.0(c), II.

12. Claims 4, 5, 8, 11, 15, 21, 25, 56-61 are rejected as being dependent upon a rejected base claim.

13. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

14. Claims 7, 12, and 20 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The newly added limitation of "executed by the client-server hardware architecture "to each "means for" causes the claim to be unclear. The claim falls within the scope of 112, sixth paragraph, where "means for" language is used, the corresponding function is disclosed, and sufficient structure is not disclosed. However, there is no disclosure of the structure for performing the recited functions. The limitations in each paragraph, as amended, are directed towards software, which is used, to perform the intended function. For example, paragraphs [0101], [0156], and

[0212]-[0213], are cited to disclose the tokenizer (part of the fact extraction tool) tag uncrossing tool, and RUBIE pattern match tool that is used to perform the intended function, which are all software components (see [0184], FEX runs). However, the interaction between such software and structural component is not found within the Specification to enable one to determine that such structural component is interacting with software to perform each function. The limitation "executed by the client-server architecture" is described in the specification to be an operating system environment and therefore is software with no interaction with structural components. The structural component that performs the intended functions in the claim is unclear. See MPEP 2174, III.

15. Claims 4, 5, 8, 11, 15, 21, 25, 56-61 are rejected as being dependent upon a rejected base claim.

Claim Rejections - 35 USC § 101

16. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 7, 12, 20 are rejected under 35 U.S.C. 101 because the claims are directed to a software embodiment. The claims are directed towards an application (i.e. software) as stated in the published application, paragraph [0516], where the FEX tool set exists part of a larger application. In paragraphs [0518]-[0521], the client-server hardware architecture is described to be nothing more than an operating system (i.e.

program) although the terminology of hardware is being used. Thus, the claim is directed to software and is thus non-statutory.

17. Claims 4, 5, 8, 11, 15, 21, 25, 56-61 are rejected as being dependent upon a rejected base claim.

Claim Rejections - 35 USC § 103

18. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

19. Claims 4, 7, 8, 12, 15, 41, 43, 48, 49, 52, and 56-62 are rejected under 35 U.S.C. 103(a) as being unpatentable over Collard et al. ("An XML-Based Lightweight C++ Fact Extractor") in view of Cunningham *et al.* (Developing Language Processing Component with GATE (a User Guide), 2001-2002) in view of Simov ("Building a linguistically Interpreted Corpus of Bulgarian: the BulTReeBank") in view of Krauthammer et al. ("Representing semantic information in a linear string of text using XML", 2002), hereinafter, Krauthammer.

As to claims 7 and 41, Collard teaches a fact extraction tool set for extracting information from a document, wherein the document includes text (see sect. 4.1, XML document), comprising:

means for extracting facts (see sect. 4.4, right column, 3rd full paragraph, execution of XPath statements using XPath tool) from the annotated text (see

sect. 4.1, XPath expressions used to extract facts from XML) using text pattern recognition rules (see sect. 4.1, XPath Query language and see sect. 5.6, where the XPath can include regular expression matching and string matching), wherein each text pattern recognition rule comprises a pattern that describes text of interest (see sect. 4.4, 2nd paragraph, XPath expression defines a pattern and wherein the text pattern recognition rules use regular expression-based functionality, tree-based traversal functionality based on a language that can navigate XML representations of text (see sect. 4.4, XPath expression used to find all functions at tope level of a XML document (XML tree) and see sect. 5.6, where the XPath uses string matching and can be combined with regular expression matching).

However, Collard does not specifically teach the following but Cunningham does teach, where Cunningham *et al.* teaches

a client-server hardware architecture that executes the means (see Page 93, 1st paragraph and Figure 6.1, a distributed in a IE system, and see page 54, sect. 3.3 and three bullets where the system is a framework as a backplane into which plug beans-based Creole components, user gives list of URLS and components loaded by system, which describe a client-server architecture).

means for breaking the text tokens (see sect 6.1, page 94, 1st paragraph, tokenizer)

a plurality of independent means for annotating text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph, i.e. POS, semantic tagger) with

token attributes (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type), constituent attributes (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text.), links (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal co reference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")) using XML as a basis for representing the annotated text (see page 60, sect. 4.4.1, 1st paragraph).

wherein the pattern recognition rule comprises a pattern that describes the text of interest (see page 82, 3rd paragraph, and rule below) (e.g. From the cited portion a definition of GazLocation is given for a portion of the pattern. This is an example of a rule.), a label that names the pattern for testing and debugging purposes (see page 81, 2nd paragraph and 2nd bullet) (e.g. A label; for debugging can be set in order to see any conflicts.); and an action that indicates what should be done in response to a successful matching of the pattern (see page 142, numeral2, subnumeral 2) (e.g. The algorithm in the cited section is used in the JAPE rules, which is a finite state machine and action executed), and wherein the text pattern recognition rules use regular expression based

functionality (see page 7, sect. 1.3.3., last two lines), and user defined matching functions.

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard with the annotation and use of regular expression based functionality as taught by Cunningham the purpose of extracting certain information in many languages (See Cunningham, page 92, 1st two lines).

However, Collard in view of Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching functions.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard in view of Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of extracting certain information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8), which would benefit the fact extraction as taught by Collard by allowing

user preferred edits to XPath expressions for extracting information according to user needs.

However, Collard in view of Cunningham in view of Simov do not specifically teach the resolving of conflicting annotation boundaries.

Krauthammer does teach the resolving of annotation boundaries (see page 5, left column, lines 10-14, linearized representation was used to overcome overlapping portions).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard in view of Cunningham in view of Simov with the resolving of annotation boundaries as taught by Krauthammer for the purpose of preventing invalid nesting of elements in XML and present a well-formed representation (see Krauthammer, page 5, left column, entire paragraph).

As to claim 4, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer teach al of the limitations as in claim 3, above.

Furthermore, Cunningham *et al.* teaches wherein, the attributes include tokenization (see sect. 6.1), text normalization (see , part of speech tags (see sect. 6.4.), sentence boundaries (see sect. 6.3), parse trees (see page 62, sect. 4.42, last paragraph-page 63, first three lines) (e.g. It is seen that annotations can be represented in hierarchical representation of a parse tree), semantic

attribute tagging (see. sect. 6.5) and other interesting attributes of the text (see sect. 6.6).

As to claims 8 and 48, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer teach al of the limitations as in claim 3, above.

Furthermore, Cunningham *et al.* teaches wherein,

the token attributes have a one-per-base-token alignment, where for the attribute type represented, there is an attempt to assign an attribute to each base token (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type).;

the constituent attributes are assigned yes-no values, where the entire pattern of each base token is considered to be a single constituent with respect to some annotation value (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text. Further, it is seen that the all of the tokens represent a pattern associated with the sentence.);

the links assign common identifiers to coreferring and other related patterns of base tokens (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal

coreference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")).

As to claim 12, Collard teaches a fact extraction tool set for extracting information from a document, wherein the document includes text (see sect. 4.1, XML document), comprising:

means for identifying and extracting facts (see sect. 4.4, right column, 3rd full paragraph, execution of XPath statements using XPath tool) from the annotated text (see sect. 4.1, XPath expressions used to extract facts from XML and therefore identifies) using text pattern recognition rules (see sect. 4.1, XPath Query language and see sect. 5.6, where the XPath can include regular expression matching and string matching), wherein each text pattern recognition rule comprises a pattern that describes text of interest (see sect. 4.4, 2nd paragraph, XPath expression defines a pattern and wherein the text pattern recognition rules use regular expression-based functionality, tree-based traversal functionality based on a language that can navigate XML representations of text (see sect. 4.4, XPath expression used to find all functions at tope level of a XML document (XML tree) and see sect. 5.6, where the XPath uses string matching and can be combined with regular expression matching).

However, Collard does not specifically teach the following but Cunningham does teach, where Cunningham *et al.* teaches

a client-server hardware architecture that executes the means (see Page 93, 1st paragraph and Figure 6.1, a distributed in a IE system, and see page 54, sect. 3.3 and three bullets where the system is a framework as a backplane into which plug beans-based Creole components, user gives list of URLS and components loaded by system, which describe a client-server architecture).

means for breaking the text tokens (see sect 6.1, page 94, 1st paragraph, tokenizer)

a plurality of independent means for annotating text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph, i.e. POS, semantic tagger) with token attributes (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type), constituent attributes (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text.), links (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal co reference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")) using XML as a basis for representing the annotated text (see page 60, sect. 4.4.1, 1st paragraph)

means for associating all annotations assigned to a particular piece of text (see page 81, 2nd paragraph, three bullets) (e.g. From the cited section it is

evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment), with the base tokens for that text to generate aligned annotations (e.g. This occurs when matching patterns.)

means for identifying and extracting potentially interesting pieces of information (see page 104, 6.8) in the aligned annotations by finding patterns in the attributes of the annotated text using text pattern recognition rules written in a rule based information extraction language, (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used, where the language is XML (see page 60, sect. 4.4.1, 1st paragraph), wherein the pattern recognition rule comprises a pattern that describes the text of interest (see page 82, 3rd paragraph, and rule below) (e.g. From the cited portion a definition of GazLocation is given for a portion of the pattern. This is an example of a rule.), a label that names the pattern for testing and debugging purposes (see page 81, 2nd paragraph and 2nd bullet) (e.g. A label; for debugging can be set in order to see any conflicts.); and an action that indicates what should be done in response to a successful matching of the pattern (see page 142, numeral2, subnumeral 2) (e.g. The algorithm in the cited section is used in the JAPE rules, which is a finite state machine and action executed), and wherein the text pattern recognition rules use regular expression based functionality (see page 7, sect. 1.3.3., last

two lines), and each text pattern recognition rule queries for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted (see page 81, last two paragraphs, and pages 82 and 83) (e.g. It is evident that from the input, attributes or annotations are specified and the latter citation is shown as a variety of data formats are possible and are looked upon in an existing list, which are compared (queried)).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard with the annotation and use of regular expression based functionality as taught by Cunningham the purpose of extracting certain information in many languages (See Cunningham, page 92, 1st two lines).

However, Collard in view of Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching functions.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard in view of Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of

extracting certain information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8), which would benefit the fact extraction as taught by Collard by allowing user preferred edits to XPath expressions for extracting information according to user needs.

However, Cunningham in view of Simov do not specifically teach the resolving of conflicting annotation boundaries.

Krauthammer does teach the resolving of annotation boundaries (see page 5, left column, lines 10-14, linearized representation was used to overcome overlapping portions).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard in view of Cunningham in view of Simov with the resolving of annotation boundaries as taught by Krauthammer for the purpose of preventing invalid nesting of elements in XML and present a well-formed representation (see Krauthammer, page 5, left column, entire paragraph).

As to claims 15, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 12, above.

Furthermore, Simov does teach user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

Furthermore, Cunningham *et al.* teaches editing function to name (see page 37, sect. 2.14.2, allows name to be edited of annotations, which define the pattern to detect. The annotations are used to detect various structures in the document for extraction) and define a fragment of a pattern (see page 84, 1st and 2nd paragraph) (e.g. The label is assigned to the year based on the pattern of word in or by found in the text).

As to claim 43, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 41, above.

Furthermore, Cunningham *et al.* teaches wherein in the annotating step the attributes include, orthographic (see sect 6.1, page 94, 1st paragraph), syntactic (see sect. 4.4.2, last paragraph), semantic (see sect. 6.5), pragmatic (see sect. 6.7.1, 1st paragraph) (e.g. The applicant refers to pragmatic as being identifying quotations, see Applicants specification, page 23, line 4) and dictionary-based attributes (see sect. 6.6.2 and see 6.2) (e.g. A table is used to determine id strings are of the same entity and the latter citation refers to names and cities).

As to claim 49, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 41, above.

Furthermore, Cunningham *et al.* teaches wherein the means for annotating a text further comprises means for associating all annotations

assigned to a particular piece of text (see page 81, 2nd paragraph, three bullets) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment), with the base tokens for that text to generate aligned annotations (e.g. This occurs when matching patterns.).

As to claim 52, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 41, above.

Furthermore, Cunningham *et al.* teaches wherein, text pattern recognition rule (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used) queries for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted (see page 81, last two paragraphs, and pages 82 and 83) (e.g. It is evident that from the input, attributes or annotations are specified and the latter citation is shown as a variety of data formats are possible and are looked upon in an existing list, which are compared (queried)).

As to claim 56, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, above.

Furthermore, Cunningham teaches wherein the text pattern recognition rules (see page 81, 2nd paragraph, three bullets and sect. 6.1.1) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment. LHS and RHS rules are used, where the language is XML (see page 60, sect. 4.4.1, 1st paragraph) query for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted (see page 81, last two paragraphs, and pages 82 and 83) (e.g. It is evident that from the input, attributes or annotations are specified and the latter citation is shown as a variety of data formats are possible and are looked upon in an existing list, which are compared (queried)),

As to claims 57-59, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, 12, and 41, above.

Furthermore, Cunningham et al. teaches annotation of text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph).

Furthermore, Simov teaches wherein representation of a single view of the document expressed as inline XML (see page 6, right column, code in between 3rd paragraph, where <s> and code in between<s/>, shows the single view with annotations of parts of speech in a single view).

As to claims 60-62, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, 12, and 41, above.

Furthermore, Collard teaches wherein the means for extracting uses XPath for traversing XML-based tree representation in the annotated text (see sect. 4.4, querying using XPath 3rd full paragraph, where the XPath expression is used to extract information by starting at top and looking at any level in the XML document tree).

20. Claims 5 and 45 are rejected under 35 U.S.C. 103(a) as being unpatentable over Collard in view of Cunningham *et al.* (Developing Language Processing Component with GATE (a User Guide), 2001-2002) in view of Simov et al. ("Building a Linguistically Interpreted Corpus of Bulgarian: the BulTreeBank", 2002), hereinafter, Simov in view of Krauthammer et al. ("Representing semantic information in a linear string of text using XML", 2002), hereinafter, Krauthammer as applied to claim 7, above and further in view of Cunningham et al. ("Gate: an architecture for development of robust HLT applications", 2002), hereinafter, Cunningham (2).

As to claim 5 and 45, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 7, above.

Furthermore, Cunningham teaches wherein the means for annotating the text (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph) comprises, a plurality of independent annotators, wherein each of the annotators has at least one specific annotation function (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2,

last paragraph, such as POS and semantic taggers, where each function independent of each other).

However, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer do not specifically teach the user-implemented means for specifying which of the annotators to use an the order of their use.

Cunningham (2) does teach user-implemented means for specifying which of the annotators to use an the order of their use (see Figure 1, right hand-pane, where checkboxes are used for selecting annotations, and page 3, left column, 1st full paragraph, where a GUI is used for user to specify order and which processing resources to use for a specific application.)

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard in view of Cunningham in view of Simov in view of Krauthammer with user implemented means as taught by Cunningham (2) as the system of Cunningham (2) is a more detailed view of the Gate system as in Cunningham hence the purpose of combining allows information extraction (see Cunningham (2), Abstract).

21. Claims 20, 21, and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cunningham *et al.* (Developing Language Processing Component with GATE (a User Guide), 2001-2002) in view of Simov et al. ("Building a Linguistically Interpreted Corpus of Bulgarian: the BulTreeBank", 2002), hereinafter, Simov in view of

Krauthammer et al. ("Representing semantic information in a linear string of text using XML", 2002), hereinafter, Krauthammer in view of Cunningham et al. ("Gate: an architecture for development of robust HLT applications", 2002), hereinafter, Cunningham (2).

As to claim 20, Cunningham *et al.* teaches wherein, a text annotation tool comprising:

a client-server hardware architecture that executes the means (see Page 93, 1st paragraph and Figure 6.1, a distributed in a IE system, and see page 54, sect. 3.3 and three bullets where the system is a framework as a backplane into which plug beans-based Creole components, user gives list of URLS and components loaded by system, which describe a client-server architecture).

means for breaking the text passage into its base tokens (see sect 6.1, page 94, 1st paragraph, tokenizer);

a plurality of independent annotators for annotating text with (see sect. 6.4 and 6.5, and page 62, sect. 4.4.2, last paragraph, such as POS and semantic taggers, where each function independent of each other)with token attributes (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type), constituent attributes (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text.), links (see sect. 6.6) (e.g. In this

cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal co reference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")) using XML as a basis for representing the annotated text (see page 60, sect. 4.4.1, 1st paragraph).

means for associating all annotations assigned to a particular piece of text with the base tokens for that text to generate aligned annotations. (see page 81, 2nd paragraph, three bullets) (e.g. From the cited section it is evident that a pattern is specified by specifying attributes to the tokens and then specifying an annotation based upon previous assignment), with the base tokens for that text to generate aligned annotations (e.g. This is implied when matching patterns.)

However, Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching functions and tree-based functionality.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and a tree based functionality (see page 5, sect. 4.1, lines 6-8, XPATH and see page 6, left column, 1st full paragraph, last 4 lines) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of extracting certain information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8).

However, Cunningham in view of Simov do not specifically teach the resolving of conflicting annotation boundaries.

Krauthammer does teach the resolving of annotation boundaries (see page 5, left column, lines 10-14, linearized representation was used to overcome overlapping portions).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham in view of Simov with the resolving of annotation boundaries as taught by Krauthammer for the purpose of preventing invalid nesting of elements in XML and present a well-formed representation (see Krauthammer, page 5, left column, entire paragraph).

However, Cunningham *et al.* in view of Simov in view of Krauthammer do not specifically teach the user-implemented means for specifying which of the annotators to use an the order of their use.

Cunningham (2) does teach user-implemented means for specifying which of the annotators to use an the order of their use (see Figure 1, right hand-pane,

where checkboxes are used for selecting annotations, and page 3, left column, 1st full paragraph, where a GUI is used for user to specify order and which processing resources to use for a specific application.)

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Cunningham in view of Simov in view of Krauthammer with user implemented means as taught by Cunningham (2) as the system of Cunningham (2) is a more detailed view of the Gate system as in Cunningham hence the purpose of combining allows information extraction (see Cunningham (2), Abstract).

As to claim 21, Cunningham *et al.* in view of Simov teach al of the limitations as in claim 20, above.

Furthermore, Cunningham *et al.* teaches wherein, the attributes include tokenization (see sect. 6.1), text normalization (see , part of speech tags (see sect. 6.4.), sentence boundaries (see sect. 6.3), parse trees (see page 62, sect. 4.42, last paragraph-page 63, first three lines) (e.g. It is seen that annotations can be represented in hierarchical representation of a parse tree), semantic attribute tagging (see. sect. 6.5) and other interesting attributes of the text (see sect. 6.6).

As to claims 25, Cunningham *et al.* in view of Simov in view of Krauthammer teach al of the limitations as in claim 20, above.

Furthermore, Cunningham *et al.* teaches wherein,

the token attributes have a one-per-base-token alignment, where for the attribute type represented, there is an attempt to assign an attribute to each base token (see sect. 6.1, 6.1.2) (e.g. From the cited sections, once the text is broken into tokens, the attributes are identified, regarding punctuation, symbols, space, number, and orthographic type).;

the constituent attributes are assigned yes-no values, where the entire pattern of each base token is considered to be a single constituent with respect to some annotation value (see page 62, last paragraph, and page 63, 1st three lines, and table 4.1.) (e.g. From the tokenization, pos is used and tagged. Further, the annotations can be used to show the hierarchical representation of the text. Further, it is seen that the all of the tokens represent a pattern associated with the sentence.);

where the links assign common identifiers to coreferring and other related patterns of base tokens (see sect. 6.6) (e.g. In this cited section relations between identities are found for match names (see sect. 6.7) (e.g. pronominal coreference). Hence, it is implied by the reference that identifiers are used to relate associated pronouns (See page 101, "Pronoun resolution")).

22. Claims 16 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Collard in view of Cunningham *et al.* (Developing Language Processing Component with GATE (a User Guide), 2001-2002) in view of Simov et al. ("Building a

Linguistically Interpreted Corpus of Bulgarian: the BulTreeBank", 2002), hereinafter, Simov.

As to claim 16, Collard teaches a rule based information extraction language for use in identifying and extracting potentially interesting pieces of information (see sect. 4.4, 1st paragraph, used to extract facts from code) in aligned annotations in a text (see sect. 4.4, aligned annotations being XML), the language comprising a plurality of text patterns recognition rules (see sect. 5.6, 2nd paragraph, XPath uses string matching and regular expressions) that query for at least one of literal txt, attributes, and relationships (see sect 4.4, queries for function definitions in XML) found in the aligned annotations to define facts to be extracted, wherein each pattern recognition rule comprises:

a pattern that describes text of interest (see sect. 4.4, //function)

wherein text pattern recognition rules (see sect. 4.1, XPath Query language and see sect. 5.6, where the XPath can include regular expression matching and string matching) use regular expression-based functionality, tree-based traversal functionality based on a language that can navigate XML representations of text (see sect. 4.4, XPath expression used to find all functions at tope level of a XML document (XML tree) and see sect. 5.6, where the XPath uses string matching and can be combined with regular expression matching).

However, Collard does not specifically teach the following but Cunningham does teach,

a client-server hardware architecture that executes the means (see Page 93, 1st paragraph and Figure 6.1, a distributed in a IE system, and see page 54,

sect. 3.3 and three bullets where the system is a framework as a backplane into which plug beans-based Creole components, user gives list of URLS and components loaded by system, which describe a client-server architecture).

a label that names the pattern for testing and debugging purposes (see page 81, 2nd paragraph and 2nd bullet) (e.g. A label; for debugging can be set in order to see any conflicts.); and

an action that indicates what should be done in response to a matching of the pattern (see page 142, numeral2, sub numeral 2) (e.g. The algorithm in the cited section is used in the JAPE rules, which is a finite state machine and action executed), and wherein the text pattern recognition rules use regular expression based functionality (see page 7, sect. 1.3.3., last two lines); and

However, Collard in view of Cunningham *et al.* does not specifically teach annotation of tree-based attributes and user-defined matching.

Simov does teach the use of annotating with tree-based attributes (see page 4, right column, HPSG grammar processing section, converted in XML representation and see left column of page 5, tree structure) and user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction tool as taught by Collard in view of Cunningham with the tree-based attribute, tree based functionality and user defined function as taught by Simov for the purpose of

extracting certain information exceeding conditions in order to present the user with accurate information (See Simov, page 4.5, right column, 1st paragraph, lines 4-8).

As to claims 19, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer, teaches all of the limitations as in claim 12, above.

Furthermore, Simov does teach user define matching functions (see page 7, right column, sect. 4.5, 1st paragraph, where user can use tools to edit elements).

Furthermore, Cunningham *et al.* teaches editing function to name (see page 37, sect. 2.14.2, allows name to be edited of annotations, which define the pattern to detect. The annotations are used to detect various structures in the document for extraction) and define a fragment of a pattern (see page 84, 1st and 2nd paragraph) (e.g. The label is assigned to the year based on the pattern of word in or by found in the text).F.

23. Claims 11 and 51 are rejected under 35 U.S.C. 103(a) as being unpatentable over Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer as applied to claims 12 and 41, above and further in view of Marcus *et al.* ("The PENN Treebank Annotating Predicate Argument Structure", 1994).

As to claim 11 and 51, Cunnigham *et al.* in view of Simov in view of Krauthammer teach all of the limitations as in claim 12 and 41, above.

Furthermore, Cunningham *et al.* discloses wherein the means for identifying and extracting potentially interesting pieces of information performs the further function of recognizing both true left and right constituent attributes (see sect. 6.1.1 and page 81, 1st paragraph) (e.g. It is seen that a left and right attributes are recognized by the tokeniser. Further it is admitted in the Applicant's background that many pattern recognition languages have rules that process text in left to right fashion(see Applicant's Specification, page 3, lines 2-3)) and constituent attributes (see page 63, 1st paragraph).

However, Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer does not specifically disclose the identification of non-contiguous attributes.

Marcus *et al.* does disclose the identification of non-contiguous attributes (see page 117, sect. 6, 2nd paragraph and example at bottom of page 117 on right hand column) (e.g. An index number is added to the label of the original constituent and allows interpretation).

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the fact extraction taught by Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer with the identification of non-contiguous attributes taught by Marcus *et al.*. The motivation to have combined the references involves the ability to represent sentences where complements of verbs occur after a sentenial level verb (see Marcus *et al.*, page 117, sect. 6, 1st paragraph), which would benefit the fact extraction tool

taught by Collard in view of Cunningham *et al.* in view of Simov in view of Krauthammer for recognizing discontinuous constituents.

Conclusion

24. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Ravichandran et al. ("Learning Surface Text Patterns for a Question Answering System") is cited to disclose pattern learning and extraction. Litkowski ("Question Answering Using XML-Tagged Documents") is cited to disclose XPath for answer retrieval.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to PARAS SHAH whose telephone number is (571)270-1650. The examiner can normally be reached on MON.-THURS. 7:00a.m.-4:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Hudspeth can be reached on (571)272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/P. S./
Examiner, Art Unit 2626

/David R Hudspeth/
Supervisory Patent Examiner, Art Unit 2626

08/14/2009